Method 2: Bernstein’s inequality for matrices

**Theorem**

(Tropp’15) Let $X_1, \ldots, X_m$ be independent Hermitian random matrices such that: $\mathbb{E}(X_k) = 0$, $\|X_k\|_{op} \leq L$ almost surely, $k \in [m]$. 

Let $Y = \sum_{k \in [m]} X_k$, and $\nu(Y) := \|\mathbb{E}(Y^2)\|_{op} = \| \sum_{k \in [m]} \mathbb{E}X^2_k\|_{op}$.

Then for all $t > 0$, $\mathbb{P}(\lambda_1(Y) \geq t) \leq n \exp \left( \frac{-t^2}{2(\nu(Y) + Lt/3)} \right)$.

This implies for all $t > 0$: $\mathbb{P}(\|Y\|_{op} \geq t) \leq 2n \exp \left( \frac{-t^2}{2(\nu(Y) + Lt/3)} \right)$.

**Corollary**

For $W$ s.t. $W_{ij}$ independent up to symmetry, $\mathbb{E}W_{ij} = 0$, $|W_{ij}| \leq 1$, $\mathbb{E}W_{ij}^2 = O(d/n)$:

If $d \gg \log(n)$ then with high probability $\rho(W) = o(d)$. 
Stronger bounds on $\rho = \rho(W)$

**Theorem (Feige and Ofek, 2005)**

Let $A \in \mathbb{R}^{n \times n}$: symmetric matrix with entries independent up to symmetry, $A_{ij} \in [0, 1]$, and such that $\mathbb{E}(A_{ij}) \leq d/n$, where $d \leq n^{1/5}$. Then for some (universal) constant $\kappa > 0$, with high probability $\rho(A - \mathbb{E}(A)) \leq \kappa \sqrt{\max(d, \log(n))}$.

**Corollary**

For $d \gg \sqrt{\max(d, \log(n))}$, i.e. $d \gg \sqrt{\log(n)}$, with high probability $\rho(A - \mathbb{E} A) = o(d)$.
Proof of Bernstein matrix inequality

Lemma

For independent Hermitian matrices $X_k$, $k \in [m]$, and $Y = \sum_{k \in [m]} X_k$:

$$\mathbb{E} \text{Tr} e^{\theta Y} \leq \text{Tr} \exp \left( \sum_{k \in [m]} \ln \mathbb{E} e^{\theta X_k} \right)$$

Lemma

For Hermitian $X$ such that $\mathbb{E}(X) = 0$ and $\|X\| \leq L$ almost surely, then:

$$\forall \theta \in (0, 3/L), \begin{cases} 
\mathbb{E} e^{\theta X} \leq \exp \left( \frac{\theta^2/2}{1-\theta L/3} \mathbb{E} X^2 \right), \\
\ln \mathbb{E} e^{\theta X} \leq \frac{\theta^2/2}{1-\theta L/3} \mathbb{E} X^2,
\end{cases}$$

Lemma

For Hermitian $A, B$, if $A \preceq B$, then $\forall i \in [n], \lambda_i(A) \leq \lambda_i(B)$. Hence for all non-decreasing $f : \mathbb{R} \to \mathbb{R}$,

$$\text{Tr } f(A) \leq \text{Tr } f(B).$$
Application: Community Detection in the Stochastic Block Model

\( G(n, \{\alpha_i\}_{i \in [K]}, P) \), where \( \alpha_i > 0, \sum_{i \in [K]} \alpha_i = 1, \ P \in [0, 1]^{K \times K} \): multi-type version of the Erdős-Rényi random graph

- \( n \) vertices partitioned into \( K \) communities
- Type (community) of node \( i : \sigma_i \in [K], \ \sigma_i : \text{i.i.d.}, \sim \alpha \)
- Conditionally on \( \sigma[n] \), independently for each pair \( i, j \in [n] \): edge \( (i, j) \) present with probability \( P_{\sigma(i),\sigma(j)} \).

**Strong signal regime:** fixed \( K, \alpha, B \in \mathbb{R}^{K \times K}_+ \); \( P = (d/n)B \), with \( \lim_{n \to +\infty} d = +\infty \)
Spectral embedding

- Extract top two (more generally top $R$) eigenvalues $\lambda_1, \lambda_2$ of graph’s adjacency matrix $A \in \mathbb{R}^{n \times n}$ (ordered by absolute value: $|\lambda_1| \geq |\lambda_2| \geq \cdots$)
- Let $x_1, x_2 \in \mathbb{R}^n$: corresponding normalized eigenvectors
- Embed vertex $k \in [n]$ into $\mathbb{R}^2$ by letting $z_k := \sqrt{n}(x_1(k), x_2(k))$
Spectral embedding

• Extract top two (more generally top \( R \)) eigenvalues \( \lambda_1, \lambda_2 \) of graph’s adjacency matrix \( A \in \mathbb{R}^{n \times n} \) (ordered by absolute value: \(|\lambda_1| \geq |\lambda_2| \geq \cdots\))

• Let \( x_1, x_2 \in \mathbb{R}^n \): corresponding normalized eigenvectors

• Embed vertex \( k \in [n] \) into \( \mathbb{R}^2 \) by letting \( z_k := \sqrt{n}(x_1(k), x_2(k)) \)

→ based on PCA dimensionality reduction of \( A \) to dimension \( R \)
Example: spectral embedding for SBM

A case with $K = 4$ communities
Spectral embedding seems to reflect community structure
→ Why / when do spectral methods work?
Theorem

Assume communities are distinguishable, i.e. for each $k \neq \ell \in [K]$, there exists $m \in [K]$ such that $B_{km} \neq B_{\ell m}$.

Assume $\sqrt{\ln(n)} \ll d \ll n^\delta$ for some fixed $\delta \in ]0, 1[$. Let $R$: rank of matrix $B$. Then with high probability:

(i) the spectrum of $A$ consists of $R$ eigenvalues of order $\Theta(d)$ and $n - R$ eigenvalues of order $o(d)$.

(ii) $R$-dimensional spectral embedding reveals underlying communities: except for vanishing fraction of nodes $i \in [n],$

\[ \|z_i - z_j\| = \begin{cases} 
  o(1) & \text{if } \sigma(i) = \sigma(j), \\
  \Omega(1) & \text{if } \sigma(i) \neq \sigma(j)
\end{cases} \]
Theorem

Assume communities are distinguishable, i.e. for each $k \neq \ell \in [K]$, there exists $m \in [K]$ such that $B_{km} \neq B_{\ell m}$.
Assume $\sqrt{\ln(n)} \ll d \ll n^\delta$ for some fixed $\delta \in ]0, 1[$. Let $R$: rank of matrix $B$. Then with high probability:
(i) the spectrum of $A$ consists of $R$ eigenvalues of order $\Theta(d)$ and $n - R$ eigenvalues of order $o(d)$.
(ii) $R$-dimensional spectral embedding reveals underlying communities: except for vanishing fraction of nodes $i \in [n], 
\|z_i - z_j\| = \begin{cases} o(1) & \text{if } \sigma(i) = \sigma(j), \\ \Omega(1) & \text{if } \sigma(i) \neq \sigma(j) \end{cases}$

Corollary

Under these conditions any sensible clustering scheme (eg K-means properly initialized) correctly classifies all but vanishing fraction of nodes.
Proof strategy

\[ A = \bar{A} + W \]

\( \bar{A} : \) block matrix (useful “signal”)

- Write adjacency matrix as \( A = \bar{A} + W \) with \( \bar{A}_{ij} = \frac{d}{n} B_{\sigma(i), \sigma(j)} \)

- \( R \) leading eigen-elements of \( \bar{A} \) capture community structure

- Control perturbation of eigen-elements of a symmetric matrix \( \bar{A} \) by addition of symmetric matrix \( W \) in terms of spectral radius \( \rho(W) \) of noise matrix

- Prove bound on \( \rho(W) \) for random noise matrix \( W \)
Eigenstructure of $\bar{A}$

Block structure of $\bar{A} \Rightarrow \bar{A}x$ constant on each block $\Rightarrow$ eigenvectors associated to non-zero eigenvalue are block-constant.

For $t \in \mathbb{R}^K$ define $x := \phi(t) = (t_{\sigma(i)})_{i \in [n]} \in \mathbb{R}^n$.

Then $\bar{A}\phi(t) = d\phi(Mt)$, where $M_{uv} := B_{uv}\alpha_v$.

Lemma

Spectrum of $\bar{A}$:
$R$ eigen-pairs $(\lambda_u = d\mu_u, \bar{x}_u = \phi(t_u))$ where $(\mu_u, t_u)$: eigen-pairs of $M$ with $\mu_u \neq 0$;
$0$: eigenvalue with multiplicity $n - R$
Lemma

Under distinguishability hypothesis there exists $\epsilon > 0$ function of $B, \alpha$ such that for any choice of normalized leading eigenvectors $\bar{x}_1, \ldots, \bar{x}_R$, $z_i = \sqrt{n}(\bar{x}_1(i), \ldots, \bar{x}_R(i))^T$ verify

$$\sigma(i) \neq \sigma(j) \Rightarrow \|z_i - z_j\| \geq \epsilon > 0$$

Proof: Let $t_u \in \mathbb{R}^K$ be such that $\sqrt{n}x_u = \phi(t_u)$, and $\sqrt{\alpha} = \text{Diag}(\sqrt{\alpha_u})$. Then: $\{\sqrt{\alpha} t_u\}_{u \in [R]}$: orthonormal family by orthonormality of the $x_u$. $t_u$ eigenvectors of matrix $M = B\alpha$, hence $\sqrt{\alpha} t_u$: orthonormal family of eigenvectors of matrix $\sqrt{\alpha} B \sqrt{\alpha}$.

Thus $\sqrt{\alpha} B \sqrt{\alpha} = \sum_{u \in [R]} \mu_u (\sqrt{\alpha} t_u) (\sqrt{\alpha} t_u)^T$.

Equivalently: $B = \sum_{u \in [R]} \mu_u t_u t_u^T$.

Hence minimum of $\|z_i - z_j\|$ over $\sigma(i) \neq \sigma(j)$ strictly positive, for otherwise $B$ has two identical rows, i.e. distinguishability fails.
Proof

- Matrix $\overline{A}$ of rank $R$, spectral gaps $|\overline{\lambda_i} - \overline{\lambda_j}| = \Omega(d)$, $R$-dimensional spectral embedding with $\overline{x}_1, \ldots, \overline{x}_R$ separates clusters $V_k = \{i \in [n] : \sigma_i = k\}$

- Feige-Ofek: $d \gg \sqrt{\log n} \Rightarrow \rho := \rho(A - \overline{A}) \ll d$. Weyl’s inequality: $R$ eigenvalues $\lambda_i$ close to $\overline{\lambda}_i = \Theta(d)$, others of order $\rho \ll d$

- Davis-Kahane: eigenvectors $x_i$ such that $\langle x_i, \overline{x}_i \rangle = 1 - O((\rho/d)^2)$

Then $\sum_{i \in [n]} \|z_i - \overline{z}_i\|^2 = n \sum_{u \in [R]} \|x_u - \overline{x}_u\|^2 = n\theta$ with $\theta = O((\rho/d)^2) = o(1)$

Hence (Tchebitchev inequality): $|\{i : \|z_i - \overline{z}_i\| \geq \theta^{1/3}\}| \leq n\theta^{1/3} = o(n)$

Yields desired conclusion: except for vanishing fraction $\theta^{1/3}$ of nodes, spectral representatives $z_i$ $\theta^{1/3}$-close of corresponding $\overline{z}_i$, themselves clustered according to community structure
Graphon model

Let $\mathcal{X}$: compact metric space, $\pi$: probability measure on $\mathcal{X}$, $P(x, y): \mathcal{X}^2 \rightarrow [0, 1]$: continuous symmetric function

**Definition**

Graphon $G(n, \pi, P)$: $\sigma_i$ i.i.d. $\sim \pi$. Conditionally on $\sigma_n$, $(A_{i,j})_{i<j}$: independent, $\sim \text{Ber}(P(\sigma_i, \sigma_j))$.

Focus on **strong signal regime**: Graphon $G(n, \pi, (d/n)K)$ for fixed $\pi$ and kernel $K$, and **signal strength** $d \rightarrow \infty$
Operator theory background

Theorem

(Kato’66) Let $T$: linear self-adjoint operator on Hilbert space $L^2(\pi)$ where $\pi$: non-negative measure be compact, i.e. for any bounded set $C$, $T(C)$ is compact.

Then: $T$ admits discrete real spectrum $\{\lambda_k\}_{k \geq 1}$ and associated orthonormal collection of eigenfunctions $\{\psi_k\}_{k \geq 1}$ such that

$$Tf(x) = \sum_{k \geq 1} \lambda_k \psi_k(x) \int_{\mathcal{X}} \psi_k(y)f(y)\pi(dy).$$

Special case: for compact $\mathcal{X}$, continuous symmetric kernel $K: \mathcal{X}^2 \to \mathbb{R}$, associated operator $T: Tf(x) = \int_{\mathcal{X}} K(x, y)f(y)\pi(dy)$ is compact. Its eigen-elements $\{\lambda_k\}_{k \geq 1}, \{\psi_k\}_{k \geq 1}$ verify

$$K(x, y) = \sum_{k \geq 1} \lambda_k \psi_k(x)\psi_k(y),$$

$$\sum_{k \geq 1} \lambda_k^2 = \int_{\mathcal{X}^2} K(x, y)^2\pi(dx)\pi(dy).$$
Koltchinskii-Giné Theorem

Theorem

For compact $\mathcal{X}$ endowed with probability measure $\pi$, continuous symmetric kernel $K : \mathcal{X}^2 \to \mathbb{R}$, associated operator $T$, positive part of its spectrum $\{\lambda_i^+\}_{i \geq 1}$: $\lambda_1^+ \geq \lambda_2^+ \geq \cdots \geq 0$.

Let $\{\sigma_i\}_{i \geq 1}$: i.i.d., $\sim \pi$, and $M = \frac{1}{n} \left( K(\sigma_i, \sigma_j) \right)_{i, j \in [n]}$.

Let $i_0 : \lambda_{i_0-1}^+ > \lambda_{i_0}^+ = \cdots = \lambda_{i_0+d-1}^+ > \lambda_{i_0+d}^+$. ($d$-dimensional eigenspace)

Let $\lambda_{j,n}^+$: positive eigenvalues of $M$, and associated ON eigenvectors $v_{j,n}^+$.

Then for $j = i_0, \ldots, i_0 + d - 1$:

$\lambda_{j,n}^+ \to \lambda_j^+$ in probability as $n \to \infty$;

There exist orthonormal eigenfunctions $\psi_j^+$ of $T$ such that

$$\left\| v_{j,n}^+ - \frac{1}{\sqrt{n}} \left\{ \psi_j(\sigma_k) \right\}_{k \in [n]} \right\| \to 0 \text{ in probability as } n \to \infty.$$
Graphon reconstruction, strong signal regime

Theorem

Let $A \sim G(n, \pi, (d/n)K)$. Let $R \geq 1$ be fixed such that $|\lambda_{R+1}(T)| < |\lambda_R(T)|$. Let $u_1, \ldots, u_R$: orthonormal collection of eigenvectors of $A$ associated with $\lambda_1(A), \ldots, \lambda_R(A)$. Assume that $\sqrt{\log n} \ll d \ll n^\delta$ for some fixed $\delta \in (0, 1)$. Define

$$
\hat{K}_{ij} = \sum_{\ell=1}^R \lambda_\ell(A) u_\ell(i) u_\ell(j), \ i, j \in [n].
$$

Then with high probability

$$
\sum_{i,j \in [n]} \left[ \frac{n}{d} \hat{K}_{ij} - K(\sigma_i, \sigma_j) \right]^2 = o(n^2) + O(n^2 \epsilon^2_R),
$$

where $\sigma_i \in \mathcal{X}$: type of vertex $i \in [n]$, and

$$
\epsilon^2_R = \sum_{\ell > R} \lambda_\ell(K)^2.
$$
Interpretation: For fraction $1 - o(1)$ of pairs $i, j$, 
$\hat{K}_{ij} = \frac{d}{n} K(\sigma_i, \sigma_j) [o(1) + O(\epsilon_R)]$.

Proof idea: Write $A = \frac{d}{n} (K(\sigma_i, \sigma_j))_{i, j \in [n]} + W$

Feige-Ofek or Bernstein inequality: w.h.p. $\rho(W) = o(d)$.

Koltchinskii-Giné’s theorem: eigen-elements of 
$\frac{d}{n} (K(\sigma_i, \sigma_j)) \approx d\lambda(T), \{n^{-1/2} \psi_\ell(\sigma_k)\}_{k \in [n]}$

Hence $\hat{K}_{ij} \approx \frac{d}{n} \sum_{\ell \in [R]} \lambda(T) \psi_\ell(\sigma_i) \psi_\ell(\sigma_j)$.

Finally $K'(x, y) := K(x, y) - \sum_{\ell \in [R]} \lambda(T) \psi_\ell(x) \psi_\ell(y)$ verifies LLN:

$$\frac{1}{n^2} \sum_{i, j \in [n]} [K'(\sigma_i, \sigma_j)]^2 \rightarrow \sum_{\ell > R} \lambda(T)^2 = \epsilon_R^2$$
Example

Take $\mathcal{X} = [0, 1]$, $\pi = \mathcal{U}([0, 1])$, $g : \mathbb{R} \to \mathbb{R}$ 1-periodic continuous function, and $K(x, y) = g(x - y)$.

Then spectrum of $K$ determined by Fourier coefficients of $g$

Ex: for $g(x) = |x|$, $|x| \leq 1/2$: $g(x) = \frac{1}{4} + \sum_{k \geq 1} \frac{(-1)^k - 1}{\pi^2 k^2} \cos(2\pi x)$

$\rightarrow \lambda_1 = \frac{1}{4}$, $\lambda_{2k} = \lambda_{2k+1} = \frac{-1}{\pi^2(2k-1)^2}$, $k \geq 1$.

$\rightarrow$ Power-law decay of spectrum of A
Pairwise graphical models and Markov random fields

**Definition**

For undirected graph \( G := (V, E) \), finite alphabet \( \mathcal{X} \), functions \( \psi_i : \mathcal{X} \rightarrow \mathbb{R}_+ \), \( i \in V \), \( \psi_e : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+ \), \( e \in E \), the probability distribution \( \mu \) on \( \mathcal{X}^V \) defined by

\[
\mu(x) := \frac{1}{Z} \prod_{i \in V} \psi_i(x_i) \prod_{e=(i,j) \in E} \psi_{i,j}(x_i, x_j)
\]

is a **pairwise graphical model** with underlying graph \( G \).

The normalization constant \( Z = \sum_{x \in \mathcal{X}^V} \prod_{i \in V} \psi_i(x_i) \prod_{e=(i,j) \in E} \psi_{i,j}(x_i, x_j) \) is known as its partition function.

**Example**

Ising model: \( \mathcal{X} = \{-1, 1\} \), \( \psi_i(x_i) = e^{h_i x_i} \), \( \psi_{ij}(x_i, x_j) = e^{J_{ij} x_i x_j} \).

\( X_i \): spin at site \( i \); \( h_i \): external field at \( i \); \( J_{ij} \): coupling coefficient between sites \( i \) and \( j \).
For undirected graph $G := (V, E)$, finite alphabet $\mathcal{X}$, probability measure $\mu$ on $\mathcal{X}^V$, $(G, \mu)$: Markov field if there exist functions $\psi_K : \mathcal{X}^K \rightarrow \mathbb{R}_+$ indexed by cliques $K$ of $G$ and $Z$ such that
$$
\mu(x) = \frac{1}{Z} \prod_K \psi_K(x_K), x \in \mathcal{X}^V,
$$
where $x_K := (x_i)_{i \in K}$.

**Easy property:** For any pairwise graphical model $\mu$ with underlying graph $G$, $(\mu, G)$ is a Markov random field.

For subsets $A, B, C \subset V$, $C$ separates $A$ and $B$ in $G$ if any path in $G$ from $A$ to $B$ traverses $C$. We denote this $A \overset{c}{\rightarrow} G \rightarrow B$. 
Hammersley-Clifford Theorem

Theorem

For Markov field \((\mu, \mathcal{G})\), and \(A, B, C\) such that \(\xrightarrow{\mathcal{G}}\), then under \(\mu\), \(X_A\) and \(X_B\) are independent conditionally on \(X_C\).

Conversely, for any probability measure \(\mu\) on \(\mathcal{X}^\mathcal{V}\) such that \(\forall x \in \mathcal{X}^\mathcal{V}, \, \mu(x) > 0\) and for all \(A, B, C\) such that \(\xrightarrow{\mathcal{G}}\), under \(\mu\), \(X_A\) and \(X_B\) are independent conditionally on \(X_C\), then \((\mu, \mathcal{G})\) is a Markov field.
Proof (Markov field $\Rightarrow$ conditional independence)

Let $A \xrightarrow{C} B$. Denote $\bar{A}$, resp. $\bar{B}$ : nodes reachable from $A$, resp. $B$ without entering $C$. Let $A' = \bar{A} \setminus A$, $B' = \bar{B} \setminus B$.

Lemma

$\bar{A}$ and $\bar{B}$ are disjoint. Any clique $K$ of $G$ is included in $\bar{A} \cup C$ or in $\bar{B} \cup C$.

Corollary

For Markov field $(\mu, G)$ there exist functions $F$, $G$ such that for any $x \in \mathcal{X}^\gamma$, $\mu(x) = F(x_{\bar{A}\cup C})G(x_{\bar{B}\cup C})$.

Thus: $P(X_{A\cup C} = x_{A\cup C}) = \sum_{y_{A'}, y_{\bar{B}}} F(x_{A\cup C}, y_{A'})G(y_{\bar{B}}, x_C)$,

$P(X_{B\cup C} = x_{B\cup C}) = \sum_{z_{B'}, z_{\bar{A}}} F(z_{\bar{A}}, x_C)G(x_{B\cup C}, z_{B'})$,

$P(X_{A\cup B\cup C} = x_{A\cup B\cup C}) = \sum_{y_{A'}, z_{B'}} F(x_{A\cup C}, y_{A'})G(x_{B\cup C}, z_{B'})$,

$P(X_C = x_C) = \sum_{z_{\bar{A}}, y_{\bar{B}}} F(z_{\bar{A}}, x_C)G(y_{\bar{B}}, x_C)$.

$\Rightarrow P(X_{A\cup C} = x_{A\cup C})P(X_{B\cup C} = x_{B\cup C}) = P(X_{A\cup B\cup C} = x_{A\cup B\cup C})P(X_C = x_C)$
Proof (Conditional independence ⇒ Markov field)

Fix \( x^* \in \mathcal{X}^\mathcal{V} \); for \( S \subset \mathcal{V} \), note \( \phi_S(x_S) := \prod_{U \subseteq S} \mu(x_U, x^*_V \setminus U)(-1)^{|S\setminus U|} \)

Lemma

One has \( \mu(x) = \mu(x^*) \prod_{S \subseteq \mathcal{V}, S \neq \emptyset} \phi_S(x_S) \).

Proof:

\[
\prod_{S \subseteq \mathcal{V}, S \neq \emptyset} \phi_S(x_S) = \prod_{S \subseteq \mathcal{V}, S \neq \emptyset} \prod_{U \subseteq S} \mu(x_U, x^*_V \setminus U)(-1)^{|S\setminus U|} = \prod_{U \subseteq \mathcal{V}} \mu(x_U, x^*_V \setminus U)^{\kappa_U},
\]

where \( \kappa_U = \sum_{S \neq \emptyset, U \subseteq S \subseteq \mathcal{V}} (-1)^{|S\setminus U|} \).

Use \( \sum_{B \subseteq A} (-1)^{|B|} = \sum_{B \subseteq A} (-1)^{|A \setminus B|} = I_{A=\emptyset} \Rightarrow \sum_{S \subseteq \mathcal{V}} (-1)^{|S|} = 0 \) to establish:

\( \kappa_\emptyset = -1, \kappa_{\mathcal{V}} = 1, \emptyset \neq U \subsetneq \mathcal{V} \Rightarrow \kappa_U = 0 \).
Lemma

For $S$ not a clique of $G$, $\phi_X(x_S) \equiv 1$.

Proof: If $\exists j \in S : x_j = x_j^*$, then

$$\phi_S(x_S) = \prod_{U \subseteq S - j} \left( \frac{\mu(x_U, x_{\mathcal{V} \setminus U})}{\mu(x_{U+j}, x_{\mathcal{V} \setminus (U+j)})} \right)^{(-1)^{|S \setminus U|}} = 1.$$  

$S$ not a clique $\Rightarrow \exists i, j \in S : (i, j) \notin \mathcal{E}$.

$$\phi_S(x_S) = \prod_{U \subseteq S - i} \left( \frac{\mu(x_U, x_{\mathcal{V} \setminus U})}{\mu(x_{U+i}, x_{\mathcal{V} \setminus (U+i)})} \right)^{(-1)^{|S \setminus U|}}.$$  

For fixed $U \subseteq S - i$, $K := \mathcal{V} \setminus \{i, j\}$, let $y_k = \begin{cases} x_k & \text{if } k \in U, \\ x_k^* & \text{if } k \in \mathcal{V} \setminus U. \end{cases}$

Then, since $\{i\} \overset{G}{\rightarrow} \{j\}$,

$$\frac{\mu(x_U, x_{\mathcal{V} \setminus U})}{\mu(x_{U+i}, x_{\mathcal{V} \setminus (U+i)})} = \frac{\mathbb{P}(X_i = x_i^* | X_K = y_K) \mathbb{P}(X_j = y_j | X_K = y_K)}{\mathbb{P}(X_i = x_i | X_K = y_K) \mathbb{P}(X_j = y_j | X_K = y_K)} \cdots$$  

$$\cdots = \frac{\mathbb{P}(X_i = x_i^* | X_K = y_K)}{\mathbb{P}(X_i = x_i | X_K = y_K)},$$  

which does not depend on value $x_j$.

Thus $\phi_S(x_S) = \phi_S(x_{S-j}, x_j^*) = 1$. 

January 25, 2021 25 / 23